A FOREGROUND - BACKGROUND TIME SHARING QUEUE
WITH GENERAL SERVICE TIMES

C. Guimaraes
Universidade Federal do Rio de Janeiro

ABSTRACT

A foreground-background queue with Poisson arrivals and independent and identically distributed service times with a general distribution \( H(x) \) is analysed using the method of the imbedded Markov chain. The generating function of the stationary distribution of queue sizes and the Laplace transform of the stationary distribution of waiting times is obtained for two priority disciplines: preemptive resume and non-preemptive or head-of-the-line. Numerical examples indicate that when the coefficient of variation of service times is high considerable reduction of total expected queue size can be obtained over the first-come-first served discipline. The results found for two queues and preemptive resume discipline can be extended to a system with an arbitrary number of queues.

1. Introduction

Let's consider a single server queue with interrupted service in the following way: jobs arrive according to a Poisson process of density \( \lambda \) and join a first-come-first served "foreground" queue where each job is processed for a fixed amount of time of length \( Q \); if the job finishes within \( Q \), it departs. Otherwise it joins a "background" queue where it waits for further service. Jobs in the first queue may have preemptive resume or non-preemptive priority over the second queue. Each job requires a service time \( X \) with a general distribution function \( H(x) \) independent of the others and of the arrival process.

Similar models have assumed geometric service times (Kleinrock
or exponential service times (Adiri & Avi-Itzhak [3]), (Coffman [1]). It is well-known, however, that service times in computer systems have large coefficients of variation (standard deviation over mean) and are better characterized by distributions like the hyperexponential (Fife [3]).

The method of imbedded Markov chains will be used to find the stationary distributions of the queue size and waiting times.

Figure 1 A F-B Queue

2. Definitions and Notation

Let's call a "quantum service" $\chi_1$, the service time given to a job in the first queue, and $H_1(x)$ its distribution function. Then

$$P(\chi_1 \leq x) = H_1(x) = \begin{cases} H(x) & \text{if } x < Q \\ 1 & \text{if } x \geq Q \end{cases}$$

Let $\chi_2$ be the remaining service time given that a job requires more than $Q$ units of time for service and $H_2(x)$ its distribution function.

Then

$$P(\chi_2 \leq x) = H_2(x) = \frac{H(x+Q)-H(Q)}{1-H(Q)}$$

Let

$$\psi(s) = \int_0^\infty e^{-sx}dH_1(x) \quad \Re(s) \geq 0$$

$$\psi_i(s) = \int_0^\infty e^{-sx}dH_i(x) \quad (i=1,2) \quad \Re(s) \geq 0$$

$$a_r = \int_0^\infty x^r dH(x) \quad a \equiv a_1$$

$$a_r = \int_0^\infty x^r dH_1(x) \quad a \equiv a_1$$

$$b_r = \int_0^\infty x^r dH_2(x) \quad b \equiv b_1$$

Then